MBMG
Metabarcoding & Metagenomics

Metabarcoding and Metagenomics 5: 153–162
DOI 10.3897/mbmg.5.69691

Data Paper

# 18S rRNA amplicon sequence data (V1–V3) of the Bronx river estuary, New York

**Melissa R. Ingala**[1*], **Irena E. Werner**[2*], **Allison M. Fitzgerald**[3], **Eugenia Naro-Maciel**[4]

1 *Sackler Institute for Comparative Genomics, American Museum of Natural History, Central Park West at 79th Street, New York, 10024, USA*

2 *Biology Dept., College of Staten Island, City University of New York, 2800 Victory Boulevard, Staten Island, New York, 10314, USA*

3 *New Jersey City University, 2039 John F. Kennedy Boulevard, Jersey City, New Jersey, 07305-1597, USA*

4 *Liberal Studies, New York University, 724 Broadway, New York, 10003, USA*

Corresponding author: Eugenia Naro-Maciel (enmaciel@nyu.edu)

**Academic editor:** Florian Leese | **Received** 16 June 2021 | **Accepted** 18 August 2021 | **Published** 7 September 2021

## Abstract

Characterising and monitoring biological diversity to foster sustainable ecosystems is highly recommended as urban centres rapidly expand. However, much of New York City's biodiversity remains undescribed, including in the historically degraded, but recovering Bronx River Estuary. In a pilot study to identify organisms and characterise biodiversity patterns there, 18S rRNA gene amplicons (V1–V3 region), obtained from river sediments and surface waters of Hunts Point Riverside and Soundview Parks, were sequenced. Across 48 environmental samples collected over three seasons in 2015 and 2016, following quality control and contaminant removal, 2,763 Amplicon Sequence Variants (ASVs) were identified from 1,918,463 sequences. Rarefaction analysis showed sufficient sampling depth, and community composition varied over time and by substrate at the study sites over the sampling period. Protists, plants, fungi and animals, including organisms of management concern, such as Eastern oysters (*Crassostrea virginica*), wildlife pathogens and groups related to Harmful Algal Blooms, were detected. The most common taxa identified in river sediments were annelid worms, nematodes and diatoms. In the water column, the most commonly observed organisms were diatoms, algae of the phylum Cryptophyceae, ciliates and dinoflagellates. The presented dataset demonstrates the reach of 18S rRNA metabarcoding for characterising biodiversity in an urban estuary.

## Key Words

eDNA, metabarcoding, New York City, next-generation sequencing, QIIME2, urban ecology

## Introduction

With the extensive modification of ecosystems by people and increasing urbanisation, the disrupted ecology of cities is garnering substantial research interest (Alberti 2008). Understanding how urban life forms are responding to habitat alteration, invasion by non-native species, pollution, human population growth, overexploitation, disease, climate change and/or interactions amongst these threats are key areas of urban ecology research. New York City is one of the world's great metropolises, yet much of its biodiversity remains to be identified and described, especially the microfauna and microflora of its rivers and estuaries.

The once highly polluted, 23 mile-long (ca. 37 km), Bronx River is currently considered "impaired", with pollutants including faecal coliforms, garbage, refuse, polychlorinated biphenyls (PCBs) and other toxins, coming from a combination of urban and stormwater runoff, Combined Sewer Overflow (CSO) outfalls, contaminated sediments, and other sources (BRA 2021; NYSDEC 2020). These waters have been impacted by invasive species, such as green (*Carcinus maenas*) and Asian shore (*Hemigrapsus sanguineus*) crabs. They are also affected by harmful algal blooms (HABs), including *Gymnodinium* dinoflagellates (Fuss and O'Neill 2015). Eukaryotic pathogens include *Cryptosporidium parvum*

* Both authors contributed equally to this work.

(an intestinal parasite that can affect humans), and the oyster pathogens *Perkinsus marinus* and MSX (*Haplosporidium nelsoni*). However, the recovering area now hosts several clean-up and revitalisation programmes, such as targeted restoration of American eels (*Anguilla rostrata*), river herring (*Alosa pseudoharengus*, *A. aestivalis*) and eastern oysters (*Crassostrea virginica*). In the Bronx River Estuary (Fig. 1), Hunts Point Riverside Park used to be an illegal garbage disposal site, but is now an integral part of the Bronx River Greenway (Kimmelman 2012). Soundview Park waters contain a successfully-restored oyster reef identified as a key research site (Grizzle et al. 2012).

Although the first step in any ecological study is the correct identification of organisms in the focal system, there are numerous challenges in conducting biodiversity inventories (Bik et al. 2012; Bohmann et al. 2014; Taberlet et al. 2018; Deiner et al. 2021). Traditional surveys involving morphological classification and other techniques provide essential data, including abundance information. However, many taxa remain difficult to detect and describe due to a combination of their microscopic or cryptic natures, collection logistics, lack of taxonomic expertise and labour-intensive morphological assessments, leading to uncertain species identifications and biodiversity estimates. Rapid, effective and standardised approaches are needed to guide more detailed investments and cost-effectively complement morphological data for comprehensive management through improved biodiversity information (Bik et al. 2012; Bohmann et al. 2014; Rees et al. 2014; Taberlet et al. 2018; Deiner et al. 2021).

Recently developed, non-invasive and transformative environmental DNA (eDNA) metabarcoding technology provides volumes of data for biodiversity assessment via next-generation sequencing. DNA barcoding was the first worldwide effort to document life and identify species using genetic sequences from a standard DNA segment (Hebert et al. 2003). Approaches such as metabarcoding, a high-throughput extension of DNA barcoding, offer orders of magnitude more data than traditional morphological or DNA barcoding research. Multiple organisms can be identified simultaneously from genetic material extracted from environmental samples (e.g. water, air and sediment) by sequencing and analysing specific marker genes using primers that target their conserved flanking regions (Taberlet et al. 2018; Deiner et al. 2021).

Environmental DNA has numerous advantages, offers high-throughput presence, absence and relative abundance data, and can improve representation of microscopic or cryptic taxa (Bik et al. 2012; Bohmann et al. 2014; Taberlet et al. 2018; Deiner et al. 2021). Environmental DNA metabarcoding is low-impact, efficient, cost-effective, rapid and replicable. The method has been effectively used in estuaries (Chariton et al. 2010, 2015; Leray and Knowlton 2015; Taberlet et al. 2018; Afzali et al. 2021; Carraro et al. 2021; García-Machado et al. 2021) and has found previously undetected or poorly characterised organisms, in particular bacterial, protist and invertebrate taxa (Bik et al. 2012; Bohmann et al. 2014; Goldberg et al. 2015; Taberlet et al. 2018; Deiner et al. 2021; Leese et al. 2021). Laboratory, computational and data storage limitations exist and reference data for taxonomic assignment of many groups are lacking, but the methods are continuously improving (Valentini et al. 2016; Taberlet et al. 2018; Deiner et al. 2021).

One advancement that facilitates biodiversity assessment and monitoring is state-of-the-art bioinformatics pipeline development to perform quality-control and large-volume data analysis (Taberlet et al. 2018), such as for the 18S rRNA amplicon datasets presented here. Some of the advantages offered by 18S rRNA metabarcoding are broad amplification across eukaryotic kingdoms, a rapidly growing reference database due to wide marker use, as well as conservation within and divergence amongst species genetic profiles (Leray and Knowlton 2016; Taberlet et al. 2018). This marker was selected here to provide a broad overview of estuarine eukaryotic biodiversity, including microorganisms, other algae and invertebrates, that would mirror our prior 16S rRNA metabarcoding work on prokaryotes (Naro-Maciel et al. 2020). Within the 18S rRNA gene, several markers for metabarcoding are being used. For protist taxa, the V4 and the V9 regions are utilised especially often (Stoeck et al. 2009; Dunthorn et al. 2012; de Vargas et al. 2015; Boenigk et al. 2018). Here, the V1–V3 region was targeted due to the high phylogenetic resolution available using hypervariable segments V2–V4, previously demonstrated in dinoflagellates (Ki 2012) and copepods (Wu et al. 2015). Initial checks against published databases and preliminary laboratory tests supported our choice of the V1–V3 region for common taxa or those of management concern, including Eastern oysters (*Crassostrea virginica*) and HAB-related taxa. Thus, the 18S rRNA dataset, presented here, was used to identify organisms, explore biodiversity patterns and establish a baseline for future work in the Bronx River Estuary.

# Methods

## Study sites and sampling

The Bronx River Estuary was sampled from August 2015 to September 2016, monthly from May to October during low tide (Fig. 1). Samples were collected from Reach 1 (NYCParks 2021) at Hunts Point (HP, 40.82°N, 73.88°W, $n_{sediment} = 9$; $n_{water} = 8$) and Soundview (SVP, 40.81°N, 73.87°W). At Soundview, samples were obtained along a restored oyster reef (SVP-BRO: $n_{sediment} = 8$; $n_{water} = 7$) and at another estuarine site about one tenth of a mile (ca. 0.16 km) away where wild oysters were observed (SVP-BRC: $n_{sediment} = 8$; $n_{water} = 8$). To investigate two key habitats of estuarine organisms and complement ongoing conventional surveys (NYCParks 2021; Fitzgerald 2013), surface waters and benthic sediments were sampled. The former were sampled by dipping a 1-litre autoclaved jar

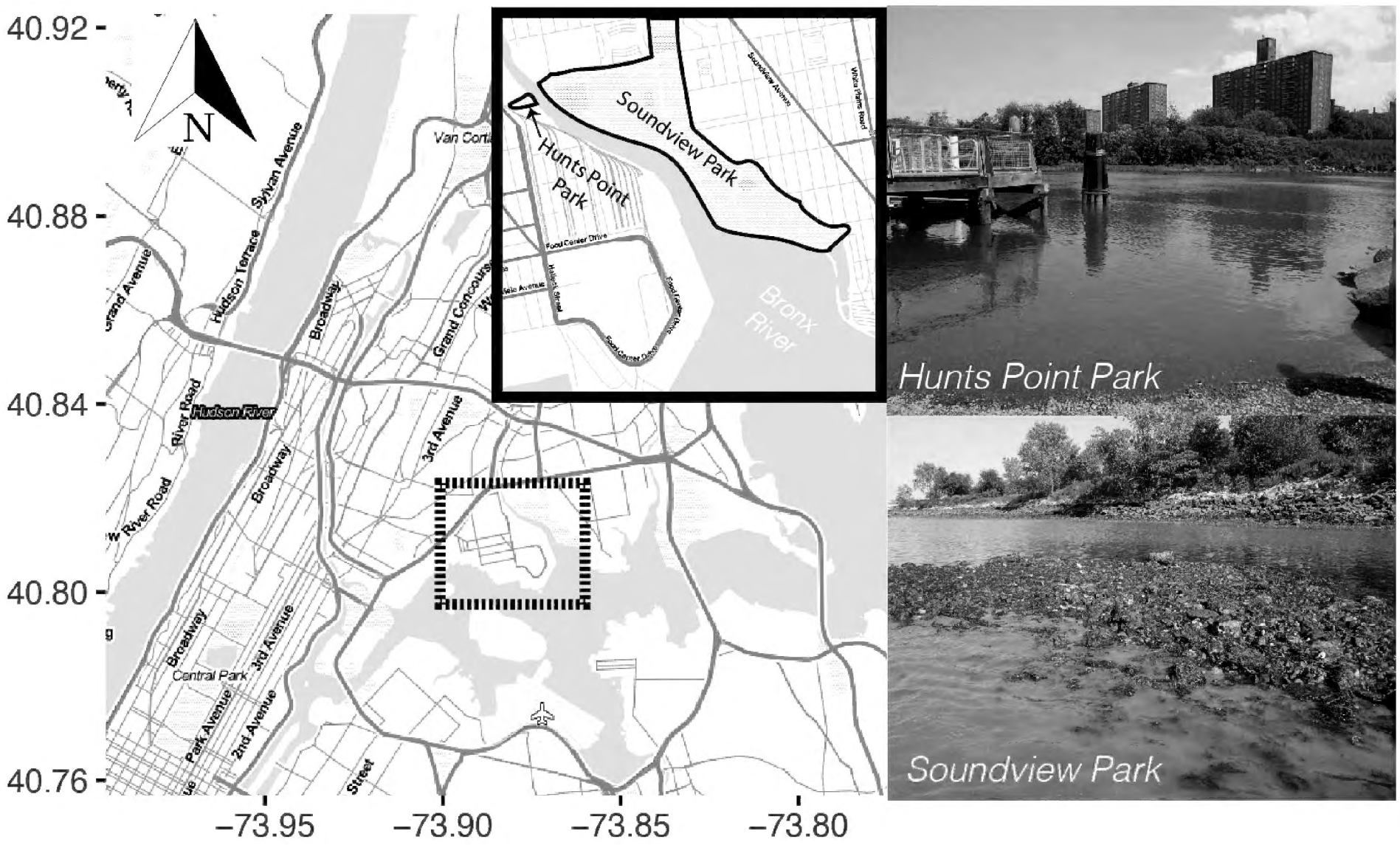


**Figure 1.** Sampling site map depicting greater New York City waterways. Inset: Detail map of sample area showing Soundview Park (SVP) and Hunts Point Riverside Park (HP) in bold outline. Right: site photos of the SVP and HP estuaries on the Bronx River. Map data 2019 (C) Google.

horizontally into the river before sediment core collection, in order to avoid contamination (Naro-Maciel et al. 2020). A polyvinyl chloride (PVC) pipe (6-inch length, 2-inch diameter) and a pallet shovel were used to sample river sediments (about 100 g) from the surface sediment layer following standard procedures, including use of disposable gloves and individual, sterilised material for each sample (Fitzgerald 2013). The samples were stored in a cooler and then moved to a laboratory refrigerator (Naro-Maciel et al. 2020).

## Environmental DNA analysis

All materials were processed within 24 hours of sampling (Naro-Maciel et al. 2020). The water samples (n = 23) were divided equally between two funnels (500 ml) and filtered with 0.45 µm Whatman Cellulose Nitrate Sterile filters (Cytiva, USA) using a standard laboratory vacuum pump (Airtech, USA; Type L-250D-G1). Filters were placed into PowerWater DNA Isolation Kit bead tubes (Qiagen, USA) and DNA was extracted following the manufacturer's instructions. A randomly selected 0.25 g soil subsample of each sediment sample (n = 25) to be used for extraction was placed in 2 ml collection tubes and centrifuged. The sediment was then transferred into PowerSoil bead tubes and extracted as instructed (Qiagen, USA). Following DNA extraction and quantification using a NanoDrop 2000 Spectrophotometer or a Qubit Fluorometer (Thermo Scientific, USA), the samples were stored frozen at -20 °C (Naro-Maciel et al. 2020). No extraction blanks or positive controls were included in this pilot study and the turtle-focused molecular biodiversity research lab was not PCR-free. However, contaminant prevention, disinfection, decontamination and sterilisation procedures, standard for a university molecular lab, were assiduously used (e.g. bleach or alcohol disinfection and surface sterilisation, single-use molecular-grade disposable material utilisation, autoclaving, UV-irradiation of supplies etc.) and state-of-the-art *in silico* quality control including contaminant identification and removal was later carried out as discussed below.

All remaining lab work (amplification, purification and sequencing) was conducted in a commercial laboratory (MRDNA, Molecular Research LP, Shallowater, TX, USA) using previously described industry-standard procedures and controls (MRDNA 2021; Dowd et al. 2008; Naro-Maciel et al. 2020). Preliminary runs with small sample sizes were conducted first to confirm primer amplification efficiency, followed by the full sample sets. Through polymerase chain reaction (PCR), about 563 bp of the 18S rRNA gene V1–V3 region was amplified using primers Euk7F (Medlin et al. 1988) and Euk570R (Weekers et al. 1994) (Euk7F: AACCTGGTTGATCCTGCCAGT + a unique 8 bp identifier barcode; Euk570R: GCTATTGGAGCTGGAATTAC). PCRs were run in 20 µl volumes using the Qiagen HotStarTaq Plus Master Mix Kit (Qiagen, USA) as follows: 94 °C for 3 minutes, 28 cycles of 94 °C for 30 seconds, 53 °C for 40 seconds and 72 °C for 1 minute and a final elongation step at 72 °C for 5 minutes (MRDNA 2021, Dowd et al.

2008). The number of cycles was determined by initial testing to optimise product detection versus errors from over-amplification.

After 2% agarose gel checks, the uniquely barcoded PCR samples were pooled in equal proportions, based on a combination of electrophoresis-based size and density estimations and DNA concentrations. The pooled samples were then purified with calibrated Ampure XP beads (Agencourt Bioscience, USA); the ratio of beads to PCR products used for purification was 0.75× as per Illumina manufacturer guidelines. Next, an Illumina DNA library was created from these purified and pooled PCR products ligated to Illumina adapters using the Illumina TruSeq DNA library preparation protocol. Finally, sequencing was performed on an Illumina MiSeq according to manufacturer's instructions, using paired-end 2 × 300 bp v.3 chemistry (MRDNA 2021, Dowd et al. 2008).

**Table 1.** Summary of sample data, including sample ID and statistics on the recovery of reads per sample after filtering, de-noising and chimeric sequence removal.

| Sample | Input | Filtered | % input passed filter | De-noised | Non-chimeric | % of input non-chimeric |
|---|---|---|---|---|---|---|
| **S.B.BRC** | 61653 | 51307 | 83.22 | 49865 | 49355 | 80.05 |
| **S.B.BRO** | 57043 | 46675 | 81.82 | 45055 | 43716 | 76.64 |
| **S.B.HP** | 41172 | 34193 | 83.05 | 32911 | 31537 | 76.60 |
| **S.C.BRC** | 65637 | 54030 | 82.32 | 52074 | 50699 | 77.24 |
| **S.C.BRO** | 53923 | 43911 | 81.43 | 42387 | 41716 | 77.36 |
| **S.C.HP** | 48441 | 40413 | 83.43 | 38983 | 37581 | 77.58 |
| **S.D.BRC** | 37670 | 31600 | 83.89 | 29892 | 29647 | 78.70 |
| **S.D.BRO** | 39984 | 32324 | 80.84 | 30829 | 30329 | 75.85 |
| **S.D.HP** | 38926 | 32241 | 82.83 | 30923 | 29822 | 76.61 |
| **S.E.BRC16** | 81853 | 67172 | 82.06 | 65692 | 64195 | 78.43 |
| **S.E.BRO16** | 64498 | 52646 | 81.62 | 51301 | 49170 | 76.23 |
| **S.E.HP16** | 50365 | 41017 | 81.44 | 39681 | 38571 | 76.58 |
| **S.F.BRC16** | 74188 | 62748 | 84.58 | 61411 | 60144 | 81.07 |
| **S.F.BRO16** | 72355 | 59701 | 82.51 | 57927 | 57472 | 79.43 |
| **S.F.HP16** | 56688 | 47187 | 83.24 | 46094 | 44870 | 79.15 |
| **S.G.BRC16** | 60173 | 50378 | 83.72 | 48470 | 47934 | 79.66 |
| **S.G.BRO16** | 59639 | 50520 | 84.71 | 49155 | 45773 | 76.75 |
| **S.G.HP16** | 62125 | 49689 | 79.98 | 48429 | 45891 | 73.87 |
| **S.H.BRC16** | 84518 | 62017 | 73.38 | 60776 | 60186 | 71.21 |
| **S.H.BRO16** | 64609 | 53593 | 82.95 | 51590 | 50524 | 78.20 |
| **S.H.HP16** | 48136 | 40675 | 84.5 | 39217 | 37346 | 77.58 |
| **S.I.BRC16** | 63314 | 53030 | 83.76 | 51647 | 51166 | 80.81 |
| **S.I.BRO16** | 54232 | 44896 | 82.79 | 43223 | 42157 | 77.73 |
| **S.I.HP16** | 51530 | 40901 | 79.37 | 39682 | 37770 | 73.30 |
| **S.J.HP16** | 55575 | 44475 | 80.03 | 43004 | 41364 | 74.43 |
| **W.B.BRC** | 66549 | 57336 | 86.16 | 49282 | 46759 | 70.26 |
| **W.B.BRO** | 66937 | 57216 | 85.48 | 49679 | 47022 | 70.25 |
| **W.B.HP** | 50568 | 43407 | 85.84 | 37591 | 35988 | 71.17 |
| **W.D.BRC** | 17978 | 15045 | 83.69 | 14346 | 14173 | 78.84 |
| **W.D.BRO** | 33475 | 27576 | 82.38 | 26316 | 25681 | 76.72 |
| **W.D.HP** | 23573 | 19891 | 84.38 | 17931 | 17902 | 75.94 |
| **W.E.BRC16** | 39716 | 34005 | 85.62 | 32661 | 31003 | 78.06 |
| **W.E.BRO16** | 37689 | 32271 | 85.62 | 30832 | 29645 | 78.66 |
| **W.E.HP16** | 39206 | 33090 | 84.4 | 31243 | 30725 | 78.37 |
| **W.F.BRC16** | 37895 | 32172 | 84.9 | 30931 | 30462 | 80.39 |
| **W.F.BRO16** | 43923 | 36032 | 82.03 | 34993 | 34027 | 77.47 |
| **W.F.HP16** | 70651 | 57008 | 80.69 | 56284 | 52129 | 73.78 |
| **W.G.BRC16** | 50138 | 42646 | 85.06 | 41774 | 38491 | 76.77 |
| **W.G.BRO16** | 47758 | 40433 | 84.66 | 39051 | 36563 | 76.56 |
| **W.G.HP16** | 54025 | 45735 | 84.66 | 44776 | 40909 | 75.72 |
| **W.H.BRC16** | 41440 | 35408 | 85.44 | 34245 | 32717 | 78.95 |
| **W.H.BRO16** | 42214 | 34834 | 82.52 | 33777 | 32871 | 77.87 |
| **W.H.HP16** | 55457 | 46005 | 82.96 | 44750 | 41631 | 75.07 |
| **W.I.BRC16** | 47134 | 39762 | 84.36 | 37335 | 35006 | 74.27 |
| **W.I.BRO16** | 48683 | 40787 | 83.78 | 38592 | 35516 | 72.95 |
| **W.I.HP16** | 46537 | 38277 | 82.25 | 37216 | 35306 | 75.87 |
| **W.J.HP16** | 41706 | 34926 | 83.74 | 33496 | 32281 | 77.40 |
| **W.J.SVP16** | 57638 | 48829 | 84.72 | 47289 | 42721 | 74.12 |
| **Totals** | 2509137 | | | | 1918463 | |

## Amplicon sequence variants (ASVs)

The FASTQ Processor was used to extract barcodes and sort forward and reverse reads into distinct files (MRDNA 2021). Raw reads were then processed using the QIIME2 v. 2019.1 pipeline (Bolyen et al. 2018) (Suppl. material 1: Document S1). The demultiplexed reads were not merged due to insufficient overlap. As quality statistics were high for forward reads, but more variable for reverse, only the forward reads were analysed as single-end. The DADA2 plug-in was then used in QIIME2 to de-noise and quality-filter the forward sequences, assign amplicon sequence variants (ASVs or features), include only Eukaryotes and generate a feature table of ASVs and metadata (Callahan et al. 2016). Primers and low-quality base calls (5–10 bp) were trimmed at the ends of each single-end sequence during the DADA2 step and reads were truncated at 260 bp, based on examination of quality scores, to account for typically observed end-of-sequence decline in quality. All other parameters, including culling short and otherwise low-quality sequences, identifying and deleting chimeras etc. were run as default (QIIME2 2021). After DADA2 filtering, the average percentage of sequences retained was 79%, with a median of 39,758 sequences kept per sample (Table 1). To taxonomically classify the 18S reads, the q2-Naïve Bayesian classifier, as implemented in QIIME2, was employed, using the SILVA 138 reference database (Quast et al. 2013) trained on the entirety of the 18S rRNA gene (Bokulich et al. 2018; Karst et al. 2018). This reference database was selected because it is a comprehensive, frequently updated and quality-curated resource for identifying eukaryotic 18S rRNA gene sequences. ASV taxonomy was manually inspected to ensure adequate taxonomic resolution was achieved.

Once taxonomic annotation was complete, R Studio v. 1.2.1335 (R Core Team 2008) was used to perform statistical analysis (Suppl. material 1: Document S2). The 'DECONTAM' programme v. 1.8.0 was run on the feature table to remove potential contaminants (Davis et al. 2018). The programme's frequency method option works by inferring potential contaminants using a simple inverse linear correlation between initial sample DNA concentration and the frequency of each ASV. Contaminants should behave such that their relative proportion increases as sample concentration decreases (Davis et al. 2018). Using a threshold of $p < 0.10$, the programme filtered ASVs meeting this criterion from the dataset. In total 28 contaminants were removed from the feature table, representing just 1% of the dataset (Suppl. material 2: Table S1).

The PHYLOSEQ v. 1.28.0 package was then used for basic data manipulation and, visualisation and community-level statistical analyses were performed using tools available in the VEGAN v. 2.5.5 package (McMurdie and Holmes 2013; Oksanen 2019). Observed ASV richness and the Shannon Diversity Index (Shannon

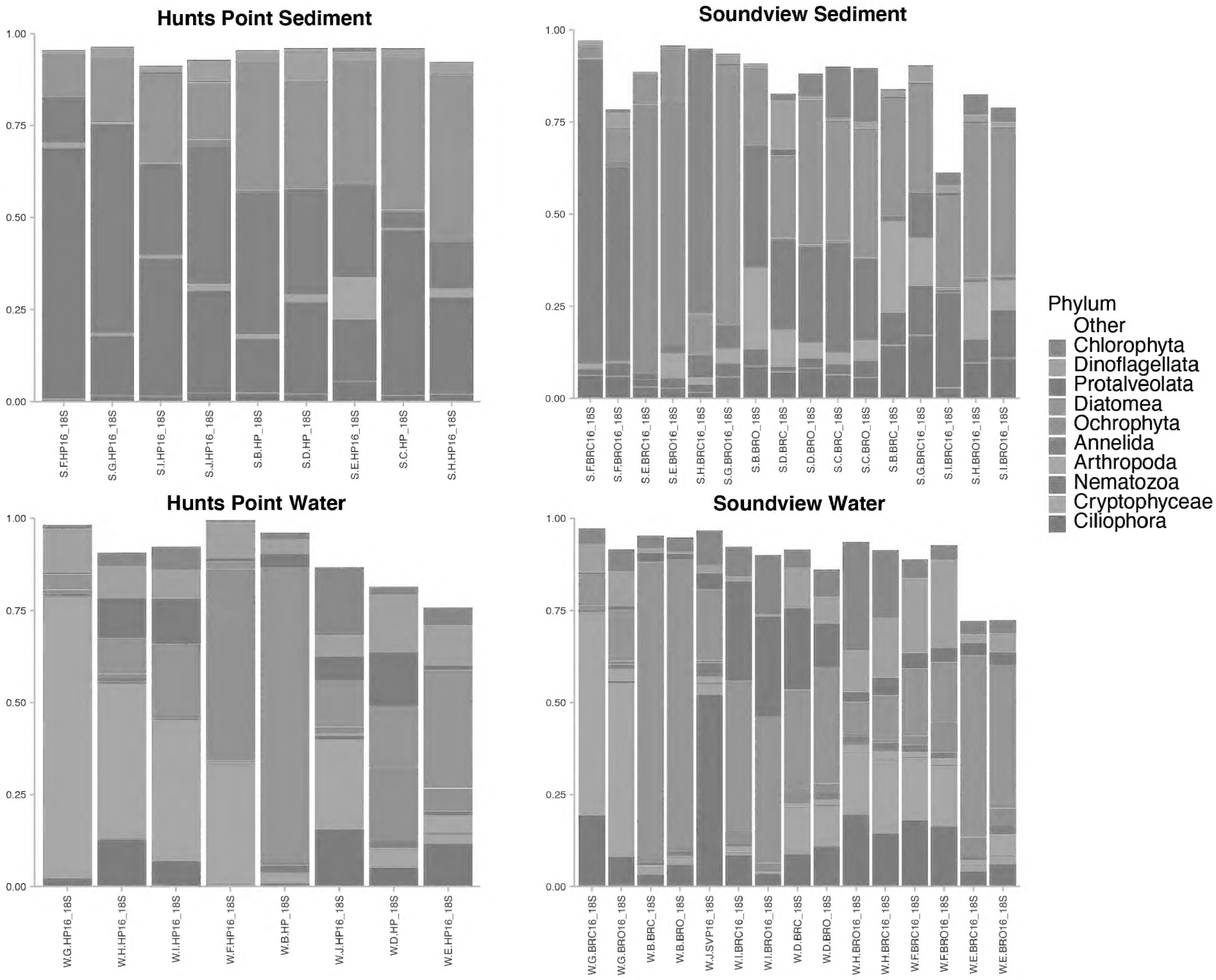


**Figure 2.** 18S rRNA community profiles of Amplicon Sequence Variants (ASVs) in sediment and water samples from Hunts Point Riverside and Soundview Parks shown at the level of phylum. Bar heights show relative abundance of sequences from each taxon.

1948) were computed to summarise alpha diversity of the eukaryotic communities. Differences in alpha diversity amongst sediments and water from both sites were assessed using a Kruskal-Wallis test with non-parametric pairwise comparisons. For beta diversity comparisons, the data were normalised using a Hellinger Transformation, which takes the square root of each ASV's relative abundance and bounds it between 0 and 1 (Legendre and Gallagher 2001; Lahti et al. 2017). Beta diversity, or turnover between sites, was summarised using the Bray-Curtis Index. To visualise differences in beta diversity, non-metric multidimensional scaling (NMDS) ordination (stress = 0.13), based on Bray-Curtis Dissimilarity was carried out using several random starts and stress assessment through the metaMDS command (Oksanen 2015) and ellipses were drawn using the stat_ellipse function.

## Results and discussion

This pilot study identified key organisms, explored biodiversity patterns and established a baseline for future work in the area, but the data must be interpreted with caution considering methodological issues. In total 48 environmental samples were successfully collected and sequenced for the 18S rRNA gene ($n_{water} = 23$; $n_{sediment} = 25$). Within these samples, protists, plants, fungi and animals encompassing 2,763 ASVs were recovered from a total of 1,918,463 post-quality-control sequences (Suppl. material 2: Table S2). Species accumulation curves of each sample reached an asymptote, indicating that the communities were surveyed with sufficient depth to detect robust differences in community structure and composition (Suppl. material 2: Fig. S1). At the study sites over the sampling period, community composition varied over time and by substrate (Fig. 2).

Several organisms of known occurrence, including taxa of management concern, were detected. Commonly observed species identified in this survey included soft-shell clams (*Mya arenaria*) and blue mussels (*Mytilus edulis*) (Suppl. material 2: Table S1). Oyster DNA (*Crassostrea virginica*) was detected in both Soundview waters and sediments, but not in Hunts Point waters or sediment. Similarly, oysters have been observed at Soundview, but not at Hunts Point; the oyster parasite genus *Perkinsus* was detected only in the water at Soundview Park. The

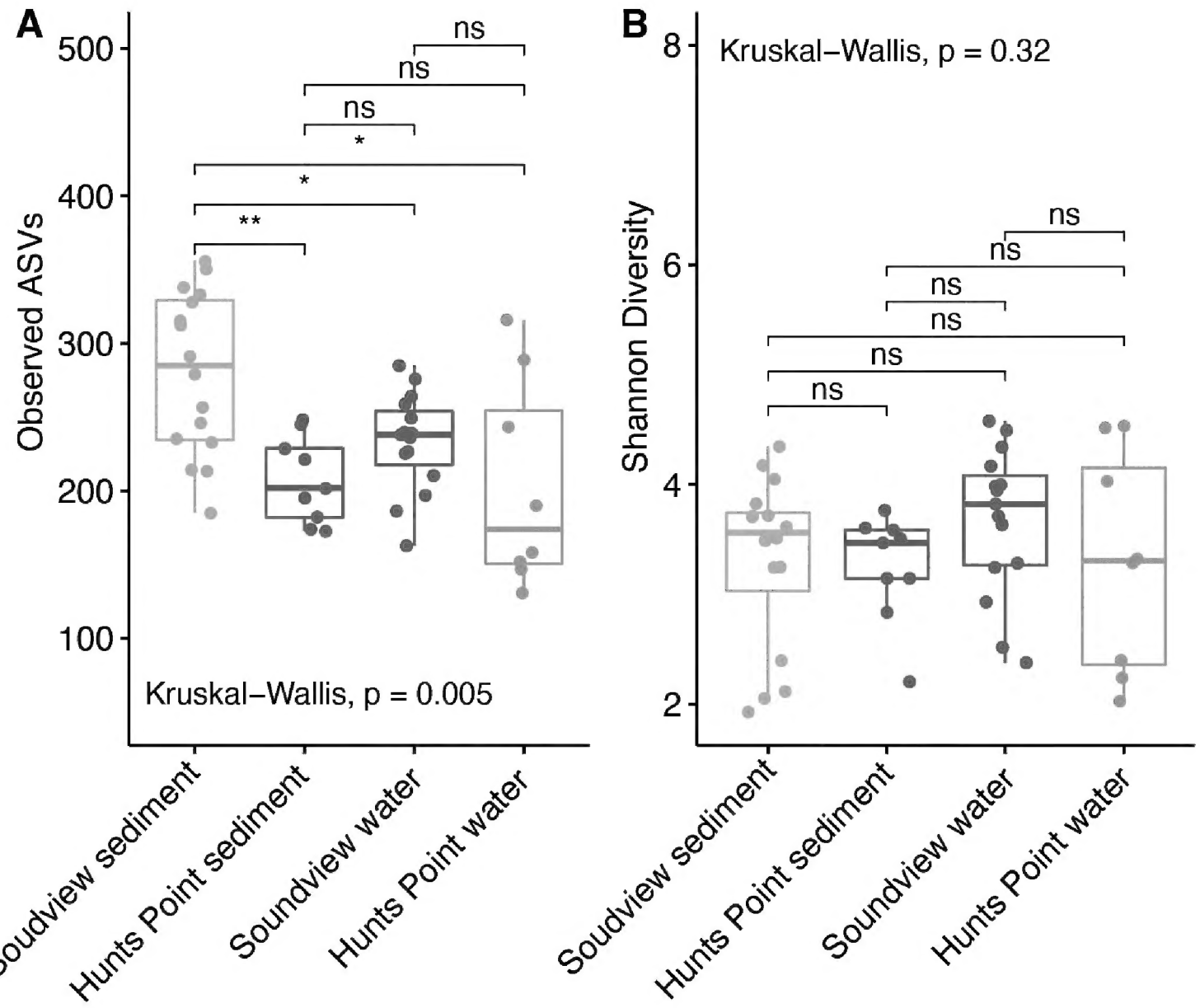


**Figure 3.** Alpha diversity comparison between sediment and water samples from Hunts Point Riverside and Soundview Parks computed using A) Observed ASVs and B) Shannon Diversity. Pairwise comparisons are indicated as follows: * = $p < 0.05$, ** = $p < 0.001$, ns = non-significant.

oyster pathogen MSX (*Haplosporidium nelsoni*) was not identified at any site, but the crustacean parasite *Haliphthoros* was found in Hunts Point sediments. Dinophyceae dinoflagellates (*Gymnodinium*, *Heterocapsa*, *Karlodinium*) and Raphidophyceae (*Heterosigma akashiwo*), which may cause Harmful Algal Blooms (Hara and Chihara 1987; Faust and Gulledge 2002; Millette et al. 2015; Lin et al. 2018), were sequenced from the Bronx River Estuary. Macroinvertebrate taxa considered to be indicators of estuary pollution (Pelletier et al. 2010; Smith et al. 2015) were not commonly found, except for the various small aquatic worms (Nematoda), some of which are consistent with poor water quality (Fig. 2). American eels (*Anguilla rostrata*) and herring (*Alosa pseudoharengus* and *A. aestivalis*), key organisms being restored and monitored in the Bronx River, were not detected.

In terms of alpha diversity, Soundview Park sediments were significantly higher in the observed number of ASVs compared with all other sites (Kruskal-Wallis, $p = 0.05$, Fig. 3A). While Soundview water also trended towards higher biodiversity compared with Hunts Point water, the difference was non-significant. However, none of the sites differed significantly in Shannon Diversity ($p > 0.05$, Fig. 3B). Sediment communities between Hunts Point and Soundview were differentiated by the presence of several key taxa missing or less proportionally abundant at Hunts Point. For example, Soundview sediments had higher proportions of arthropod DNA detected than those at Hunts Point (Figs 2 and 3). In agreement with our results on overall alpha diversity metrics, Soundview Park sediments were more taxonomically diverse when compared with Hunts Point. Water samples from both sites were not apparently different in taxonomic composition (Figs 2–4). However, there were clear differences between the taxonomic make-up of sediments and water column samples, driven mostly by the more frequent detection of annelid worms and nematodes in sediments and larger proportions of diatoms, dinoflagellates and Protalveolata in water samples (Fig. 2). The community turnover (i.e. beta diversity) of eDNA from water samples was significantly different from that of sediment ($r^2 = 0.169$, $P = 0.001$; Fig. 4). Water samples were homogeneous amongst sites ($r^2 = 0.069$, $P = 0.834$). In contrast, sediment samples from Soundview Park were distinct from those at Hunts Point ($r^2 = 0.245$, $P = 0.001$; Fig. 4).

Future metabarcoding work in the area would benefit from lessons learned during and resources developed since this pilot study. The high quality, comprehensive protocols now available to standardise and ensure eDNA metabarcoding excellence should be carefully followed (Taberlet et al. 2018; Minamoto et al. 2021). To better incorporate unicellular organisms and viruses, specialised methods, including use of filters with finer pore size, should be employed. While state-of-the art bioinformatics work conservatively identified and removed errors and contaminants, ongoing research would additionally

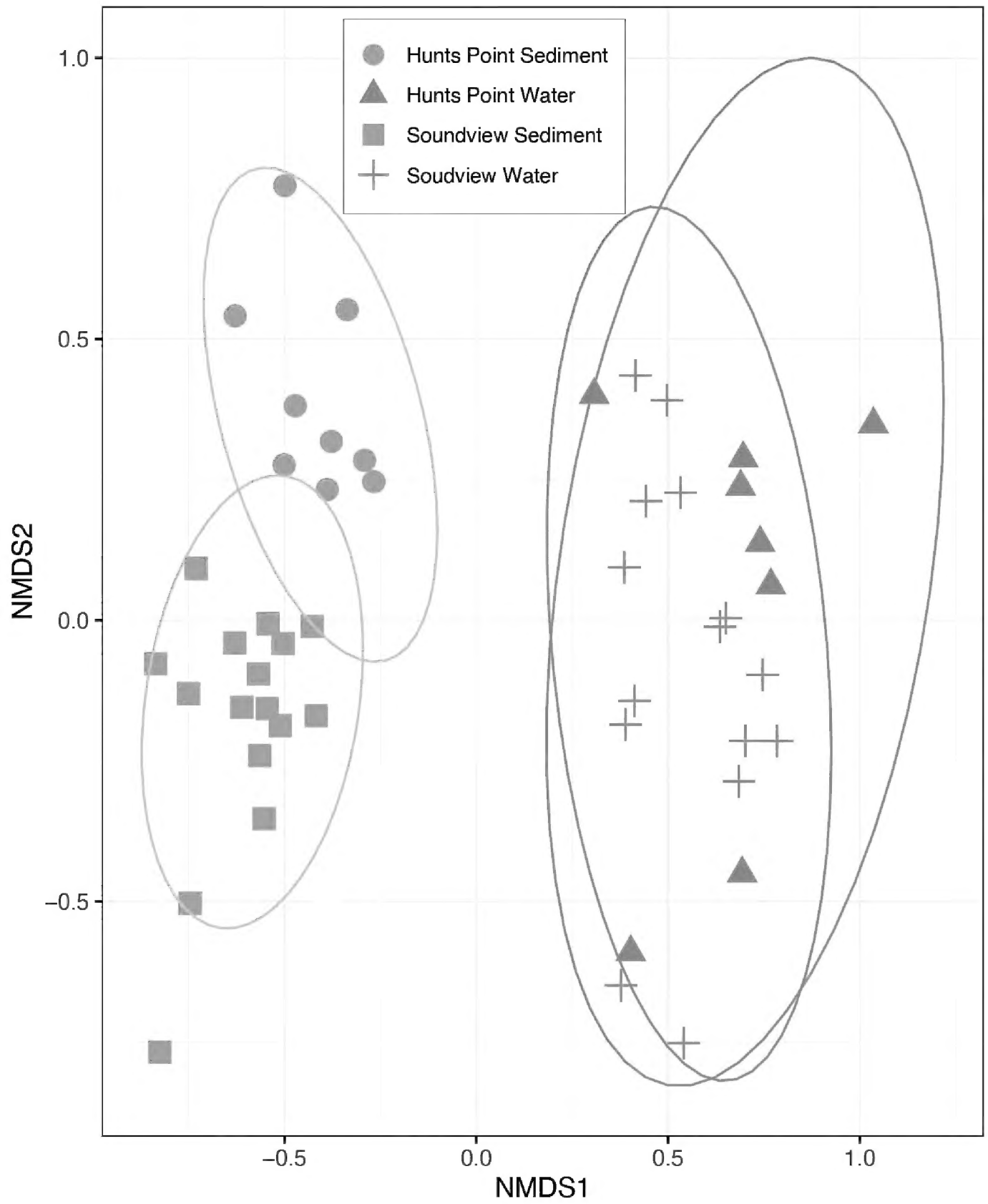


**Figure 4.** Community comparisons amongst substrates (sediment and water) and sites (Hunts Point Riverside and Soundview Parks), based on Amplicon Sequence Variants (ASVs) and using Non-metric multidimensional scaling analysis (NMDS).

benefit from stringent laboratory checks. These include the use of extraction blanks in a PCR-free lab, positive controls to assess amplification efficiency and negative controls to identify contaminants directly. Technical replicates address contamination, errors including rare taxa detection, and PCR amplification and sequencing variation. Further, although the V1 – V3 segment did capture known organisms and those of management interest, reference data bases for the increasingly used V4 or V9 regions may be more complete, thus resulting in more identifications. Finally, organismal abundance does not necessarily correlate with sequence abundance given amplification biases and errors such as primer-template fidelity and suboptimal annealing temperatures. Thus even inferences about relative abundance should be interpreted with caution (Fonseca 2018; Taberlet et al. 2018). Comparing metabarcoding results to conventional survey data will continue to be essential for ground-truthing and optimising both methods (Fediajevaite et al. 2021).

In conclusion, the 18S rRNA V1 – V3 dataset, presented here, complements our prior study, "16S rRNA Amplicon Sequencing of Urban Prokaryotic Communities in the South Bronx River Estuary". Future work will comparatively analyse information from these two genetic regions and new data from Cytochrome Oxidase I, the standard locus for animal barcoding (Hebert et al. 2003). Despite its advantages, the 18S rRNA marker alone is insufficient to fully characterise biodiversity and a suite of markers would provide a more complete profile (Leray and Knowlton 2016; Taberlet et al. 2018) to further describe life in a complex urban estuary.

## Data availability

All 18S rRNA amplicon gene sequences from this study are posted on the NCBI Sequence Read Archive (SRA) under BioProject PRJNA606795 accession numbers

SAMN19729835–SAMN19729882 (Table 1). All DNA extracts are stored at the American Museum of Natural History. Bioinformatics and statistical scripts are available as a supplement to this article (Suppl. material 1: Documents S1, S2).

## Acknowledgements

We are grateful for the New York University Research Challenge Fund and NYU Liberal Studies New Faculty Scholarship and Creative Production Awards (to ENM) and to private donors through Experiment.com (to IW) for funding the research. Site access was provided by NY/NJ Baykeeper and the New York City Department of Parks and Recreation (Natural Resources Group). Special thanks are extended to student assistants Christian Bojorquez, NaVonna Truner, Sean Thomas, Jennifer Servis, Patrick Shea, Vanessa Van Deusen and Seth Wollney, as well as to Dr .Brendan Reid.

## Conflicts of Interest

The authors declare no competing interests.

## References

Afzali SF, Bourdages H, Laporte M, Mérot C, Normandeau E, Audet C, Bernatchez L (2021) Comparing environmental metabarcoding and trawling survey of demersal fish communities in the Gulf of St. Lawrence, Canada. Environmental DNA 3: 22–42. https://doi.org/10.1002/edn3.111

Alberti M (2008) Advances in urban ecology. Springer, Seattle, 366 pp. https://doi.org/10.1007/978-0-387-75510-6

Bik HM, Porazinska DL, Creer S, Caporaso JG, Knight R, Thomas WK (2012) Sequencing our way towards understanding global eukaryotic biodiversity. Trends in Ecology & Evolution 27: 233–243. https://doi.org/10.1016/j.tree.2011.11.010

Boenigk J, Wodniok S, Bock C, Beisser D, Hempel C, Grossmann L, Lange A, Jensen M (2018) Geographic distance and mountain ranges structure freshwater protist communities on a European scale. Metabarcoding and Metagenomics 2: e21519. https://doi.org/10.3897/mbmg.2.21519

Bohmann K, Evans A, Gilbert MTP, Carvalho GR, Creer S, Knapp M, Yu DW, de Bruyn M (2014) Environmental DNA for wildlife biology and biodiversity monitoring. Trends in Ecology & Evolution 29: 358–367. https://doi.org/10.1016/j.tree.2014.04.003

Bokulich NA, Kaehler BD, Rideout JR, Dillon M, Bolyen E, Knight R, Huttley GA, Gregory Caporaso J (2018) Optimizing taxonomic classification of marker-gene amplicon sequences with QIIME 2's q2-feature-classifier plugin. Microbiome 6: e90. https://doi.org/10.1186/s40168-018-0470-z

Bolyen E, Dillon M, Bokulich N, Abnet C, Al-Ghalith G, Alexander H, Alm EJ, Arumugam M, Asnicar F, Bai Y, Bisanz JE, Bittinger K, Brejnrod A, Brislawn CJ, Brown CT, Callahan BJ, Caraballo-Rodríguez AM, Chase J, Cope E, Da Silva R, Dorrestein PC, Douglas GM, Durall DM, Duvallet C, Edwardson CF, Ernst M, Estaki M, Fouquier J, Gauglitz JM, Gibson DL, Gonzalez A, Gorlick K, Guo J, Hillmann B, Holmes S, Holste H, Huttenhower C, Huttley G, Janssen S, Jarmusch AK, Jiang L, Kaehler B, Kang KB, Keefe CR, Keim P, Kelley ST, Knights D, Koester I, Kosciolek T, Kreps J, Langille MGI, Lee J, Ley R, Liu Y-X, Loftfield E, Lozupone C, Maher M, Marotz C, Martin BD, McDonald D, McIver LJ, Melnik AV, Metcalf JL, Morgan SC, Morton J, Naimey AT, Navas-Molina JA, Nothias LF, Orchanian SB, Pearson T, Peoples SL, Petras D, Preuss ML, Pruesse E, Rasmussen LB, Rivers A, Robeson II MS, Rosenthal P, Segata N, Shaffer M, Shiffer A, Sinha R, Song SJ, Spear JR, Swafford AD, Thompson LR, Torres PJ, Trinh P, Tripathi A, Turnbaugh PJ, Ul-Hasan S, van der Hooft JJJ, Vargas F, Vázquez-Baeza Y, Vogtmann E, von Hippel M, Walters W, Wan Y, Wang M, Warren J, Weber KC, Williamson CHD, Willis AD, Xu ZZ, Zaneveld JR, Zhang Y, Zhu Q, Knight R, Caporaso G (2018) QIIME 2: Reproducible, interactive, scalable, and extensible microbiome data science. PeerJ Preprints e27295v1. https://doi.org/10.7287/peerj.preprints.27295v2

BRA (2021) BRA About The River – Bronx River Alliance. http://bronxriver.org/?pg=content&p=abouttheriver [July 9, 2021]

Callahan BJ, McMurdie PJ, Rosen MJ, Han AW, Johnson AJA, Holmes SP (2016) DADA2: High-resolution sample inference from Illumina amplicon data. Nature Methods 13: 581–583. https://doi.org/10.1038/nmeth.3869

Carraro L, Stauffer JB, Altermatt F (2021) How to design optimal eDNA sampling strategies for biomonitoring in river networks. Environmental DNA 3: 157–172. https://doi.org/10.1002/edn3.137

Chariton AA, Court LN, Hartley DM, Colloff MJ, Hardy CM (2010) Ecological assessment of estuarine sediments by pyrosequencing eukaryotic ribosomal DNA. Frontiers in Ecology and the Environment 8: 233–238. https://doi.org/10.1890/090115

Chariton AA, Stephenson S, Morgan MJ, Steven ADL, Colloff MJ, Court LN, Hardy CM (2015) Metabarcoding of benthic eukaryote communities predicts the ecological condition of estuaries. Environmental Pollution 203: 165–174. https://doi.org/10.1016/j.envpol.2015.03.047

Davis NM, Proctor DiM, Holmes SP, Relman DA, Callahan BJ (2018) Simple statistical identification and removal of contaminant sequences in marker-gene and metagenomics data. Microbiome 6: 1–14. https://doi.org/10.1186/s40168-018-0605-2

Deiner K, Yamanaka H, Bernatchez L (2021) The future of biodiversity monitoring and conservation utilizing environmental DNA. Environmental DNA 3: 3–7. https://doi.org/10.1002/edn3.178

Dowd SE, Sun Y, Wolcott RD, Domingo A, Carroll JA (2008) Bacterial Tag–Encoded FLX Amplicon Pyrosequencing (bTEFAP) for Microbiome Studies: Bacterial Diversity in the Ileum of Newly Weaned Salmonella -Infected Pigs. Foodborne Pathogens and Disease 5: 459–472. https://doi.org/10.1089/fpd.2008.0107

Dunthorn M, Klier J, Bunge J, Stoeck T (2012) Comparing the Hyper-Variable V4 and V9 Regions of the small subunit rDNA for assessment of ciliate environmental diversity. Journal of Eukaryotic Microbiology 59: 185–187. https://doi.org/10.1111/j.1550-7408.2011.00602.x

Faust MA, Gulledge RA (2002) Identifying harmful marine dinoflagellates. Contributions from the United States National Herbarium 42: 1–144. http://www.jstor.org/stable/23493225

Fediajevaite J, Priestley V, Arnold R, Savolainen V (2021) Meta-analysis shows that environmental DNA outperforms traditional surveys, but warrants better reporting standards. Ecology and Evolution 11: 4803–4815. https://doi.org/10.1002/ece3.7382

Fitzgerald AM (2013) The effects of chronic habitat degradation on the physiology and metal accumulation of eastern oysters (*Crassostrea virginica*) in the Hudson Raritan Estuary. PhD Thesis. Graduate Center, The City University of New York.

Fonseca VG (2018) Pitfalls in relative abundance estimation using eDNA metabarcoding. Molecular Ecology Resources 18: 923–926 https://doi.org/10.1111/1755-0998.12902.

Fuss, O'Neill (2015) Citizen Science on the Bronx River: An Analysis of Water Quality Data. Bronx, New York.

García-Machado E, Laporte M, Normandeau E, Hernández C, Côté G, Paradis Y, Mingelbier M, Bernatchez L (2021) Fish community shifts along a strong fluvial environmental gradient revealed by eDNA metabarcoding. Environmental DNA 00: 1–18. https://doi.org/10.1002/edn3.221

Goldberg CS, Strickler KM, Pilliod DS (2015) Moving environmental DNA methods from concept to practice for monitoring aquatic macroorganisms. Biological Conservation 183: 1–3. https://doi.org/10.1016/j.biocon.2014.11.040

Grizzle R, Ward K, Lodge J, Suszkowski D, Mosher-Smith K, Kalchmayr K, Malinowski P (2012) Oyster Restoration Research Project (ORRP) Technical Report. New York. http://www.harborestuary.org/pdf/HabitatPages/OysterReefs/ORRPFINAL_REPORT_2013_02_20.pdf

Hara Y, Chihara M (1987) Morphology, ultrastructure and taxonomy of the raphidophycean alga *Heterosigma akashiwo*. The Botanical Magazine Tokyo 100: 151–163. https://doi.org/10.1007/BF02488320

Hebert PDN, Cywinska A, Ball SL, DeWaard JR (2003) Biological identifications through DNA barcodes. Proceedings of the Royal Society B: Biological Sciences 270: 313–321. https://doi.org/10.1098/rspb.2002.2218

Karst SM, Dueholm MS, McIlroy SJ, Kirkegaard RH, Nielsen PH, Albertsen M (2018) Retrieval of a million high-quality, full-length microbial 16S and 18S rRNA gene sequences without primer bias. Nature Biotechnology 36: 190–195. https://doi.org/10.1038/nbt.4045

Ki J-S (2012) Hypervariable regions (V1–V9) of the dinoflagellate 18S rRNA using a large dataset for marker considerations. Journal of Applied Phycology 24: 1035–1043. https://doi.org/10.1007/s10811-011-9730-z

Kimmelman M (2012) Bronx River Now Flows by Parks. The New York Times. https://www.nytimes.com/2012/07/22/arts/design/bronx-river-now-flows-by-parks.html [June 4, 2021]

Lahti L, Shetty S, Blake T (2017) Tools for microbiome analysis in R. Microbiome Package Version 0.99. http://microbiome.github.com/microbiome

Leese F, Sander M, Buchner D, Elbrecht V, Haase P, Zizka VMA (2021) Improved freshwater macroinvertebrate detection from environmental DNA through minimized nontarget amplification. Environmental DNA 3: 261–276. https://doi.org/10.1002/edn3.177

Legendre P, Gallagher ED (2001) Ecologically meaningful transformations for ordination of species data. Oecologia 129: 271–280. https://doi.org/10.1007/s004420100716

Leray M, Knowlton N (2015) DNA barcoding and metabarcoding of standardized samples reveal patterns of marine benthic diversity. Proceedings of the National Academy of Sciences 112: 2076–2081. https://doi.org/10.1073/pnas.1424997112

Leray M, Knowlton N (2016) Censusing marine eukaryotic diversity in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences 371: 20150331. https://doi.org/10.1098/rstb.2015.0331

Lin C-H [Michelle], Lyubchich V, Glibert PM (2018) Time series models of decadal trends in the harmful algal species *Karlodinium veneficum* in Chesapeake Bay. Harmful Algae 73: 110–118. https://doi.org/10.1016/j.hal.2018.02.002

McMurdie PJ, Holmes S (2013) phyloseq: An R Package for Reproducible Interactive Analysis and Graphics of Microbiome Census Data. PLOS ONE 8: e61217. https://doi.org/10.1371/journal.pone.0061217

Medlin L, Elwood HJ, Stickel S, Sogin ML (1988) The characterization of enzymatically amplified eukaryotic 16S-like rRNA-coding regions. Gene 71: 491–499. https://doi.org/10.1016/0378-1119(88)90066-2

Millette N, Stoecker D, Pierson J (2015) Top-down control by micro- and mesozooplankton on winter dinoflagellate blooms of *Heterocapsa rotundata*. Aquatic Microbial Ecology 76: 15–25. https://doi.org/10.3354/ame01763

Minamoto T, Miya M, Sado T, Seino S, Doi H, Kondoh M, Nakamura K, Takahara T, Yamamoto S, Yamanaka H, Araki H, Iwasaki W, Kasai A, Masuda R, Uchii K (2021) An illustrated manual for environmental DNA research: Water sampling guidelines and experimental protocols. Environmental DNA 3: 8–13. https://doi.org/10.1002/edn3.121

MRDNA (2021) MRDNA Amplicon Sequencing. https://www.mrdnalab.com/ [July 8, 2021a]

MRDNA FASTQ Processor (2021) MRDNA FASTQ Processor. www.mrdnafreesoftware.com

Naro-Maciel E, Ingala MR, Werner IE, Fitzgerald AM (2020) 16S rRNA Amplicon Sequencing of Urban Prokaryotic Communities in the South Bronx River Estuary. Microbiology Resource Announcements 9(22): e00182-20. https://doi.org/10.1128/MRA.00182-20

NYCParks (2021) NYCParks Estuary Section: Wetlands of the Bronx River Watershed. https://www.nycgovparks.org/greening/natural-resources-group/bronx-river-wetlands/estuary-section [June 4, 2021]

NYSDEC (2020) NYS Section 303(d) List of Impaired/TMDL Waters – NYS Dept. of Environmental Conservation. https://www.dec.ny.gov/chemical/31290.html [June 9, 2021]

Oksanen J (2015) Multivariate analysis of ecological communities in R: vegan tutorial. http://cc.oulu.fi/~jarioksa/opetus/metodi/vegantutor.pdf [Accessed July 18, 2018. : 1–43]

Oksanen J (2019) Vegan: ecological diversity. https://cran.r-project.org/package=vegan

Pelletier MC, Gold AJ, Heltshe JF, Buffum HW (2010) A method to identify estuarine macroinvertebrate pollution indicator species in the Virginian Biogeographic Province. Ecological Indicators 10: 1037–1048. https://doi.org/10.1016/j.ecolind.2010.03.005

QIIME2 (2021) QIIME2 denoise-single: Denoise and dereplicate single-end sequences. https://docs.qiime2.org/2021.4/plugins/available/dada2/denoise-single/ [July 19, 2021]

Quast C, Pruesse E, Yilmaz P, Gerken J, Schweer T, Yarza P, Peplies J, Glöckner FO (2013) The SILVA ribosomal RNA gene database project: improved data processing and web-based tools. Nucleic Acids Research 41: D590–D596. https://doi.org/10.1093/nar/gks1219

R Core Team (2008) 739 Computational Many-Particle Physics. Fehske H, Schneider R, Weiße A (Eds). Springer Berlin Heidelberg, Berlin, Heidelberg, 780 pp. https://doi.org/10.1007/978-3-540-74686-7

Rees HC, Maddison BC, Middleditch DJ, Patmore JRM, Gough KC (2014) The detection of aquatic animal species using environmental DNA – a review of eDNA as a survey tool in ecology. Journal of Applied Ecology 51: 1450–1459. https://doi.org/10.1111/1365-2664.12306

Shannon C (1948) A mathematical theory of communication. The Bell System Technical Journal 27: 379–423. https://doi.org/10.1002/j.1538-7305.1948.tb01338.x

Smith AJ, Rickard S, Mosher EA, Lojpersberger JL, Heitzman DL, Duffy BT, Novak MA (2015) Biological stream assessment Bronx River. Albany, NY

Stoeck T, Behnke A, Christen R, Amaral-Zettler L, Rodriguez-Mora MJ, Chistoserdov A, Orsi W, Edgcomb VP (2009) Massively parallel tag sequencing reveals the complexity of anaerobic marine protistan communities. BMC Biology 7: e72. https://doi.org/10.1186/1741-7007-7-72

Taberlet P, Bonin A, Zinger L, Coissac E (2018) 1 Environmental DNA: For Biodiversity Research and Monitoring. Oxford University Press, London, 253 pp. https://doi.org/10.1093/oso/9780198767220.001.0001

Valentini A, Taberlet P, Miaud C, Civade R, Herder J, Thomsen PF, Bellemain E, Besnard A, Coissac E, Boyer F, Gaboriaud C, Jean P, Poulet N, Roset N, Copp GH, Geniez P, Pont D, Argillier C, Baudoin J-M, Peroux T, Crivelli AJ, Olivier A, Acqueberge M, Le Brun M, Møller PR, Willerslev E, Dejean T (2016) Next-generation monitoring of aquatic biodiversity using environmental DNA metabarcoding. Molecular Ecology 25: 929–942. https://doi.org/10.1111/mec.13428

de Vargas C, Audic S, Henry N, Decelle J, Mahe F, Logares R, Lara E, Berney C, Le Bescot N, Probert I, Carmichael M, Poulain J, Romac S, Colin S, Aury J-M, Bittner L, Chaffron S, Dunthorn M, Engelen S, Flegontova O, Guidi L, Horak A, Jaillon O, Lima-Mendez G, Luke J, Malviya S, Morard R, Mulot M, Scalco E, Siano R, Vincent F, Zingone A, Dimier C, Picheral M, Searson S, Kandels-Lewis S, Acinas SG, Bork P, Bowler C, Gorsky G, Grimsley N, Hingamp P, Iudicone D, Not F, Ogata H, Pesant S, Raes J, Sieracki ME, Speich S, Stemmann L, Sunagawa S, Weissenbach J, Wincker P, Karsenti E, Boss E, Follows M, Karp-Boss L, Krzic U, Reynaud EG, Sardet C, Sullivan MB, Velayoudon D (2015) Eukaryotic plankton diversity in the sunlit ocean. Science 348: 1261605–1261605. https://doi.org/10.1126/science.1261605

Weekers P, Gast R, Fuerst P, Byers T (1994) Sequence variations in small-subunit ribosomal RNAs of Hartmannella vermiformis and their phylogenetic implications. Molecular Biology and Evolution 11: 684–690. https://doi.org/10.1093/oxfordjournals.molbev.a040147

Wu S, Xiong J, Yu Y (2015) Taxonomic Resolutions Based on 18S rRNA Genes: A Case Study of Subclass Copepoda. Fontaneto D (Ed.). PLoS ONE 10: e0131498. https://doi.org/10.1371/journal.pone.0131498

**Supplementary material 1**
**Scripts used for metabarcoding analysis**
Author: Melissa Ingala
Data type: QIIME and R Scripts (in zip. archive)
Explanation note: **Document S1.** QIIME2 workflow. **Document S2.** R script.

Link: https://doi.org/10.3897/mbmg.5.69691.suppl1

**Supplementary material 2**
**Figure S1, Tables S1, S2**
Author: Melissa Ingala
Data type: jpg. file, docx. file and xslx. file (in zip. archive)
Explanation note: **Figure S1.** Sample-based species accumulation curves of total 18S rRNA diversity by substrate type (sediment, water) for each site (Hunts Point Riverside and Soundview Parks), calculated using the Vegan 2.4-3 package for Amplicon Sequence Variants (ASVs). **Table S1.** Taxonomies of contaminating ASVs removed by "decontam" analysis. **Table S2.** Taxonomic Assignment including ASV ID to Domain (d), Phylum (p), Class (c), Order (o), Family (f), Genus (g) and Species (s), as applicable.

Link: https://doi.org/10.3897/mbmg.5.69691.suppl2